New Approach for *In Silico* Genotoxicity Testing of Impurities and Degradants

Kiril Lanevskij[1,2], Liutauras Juska[1,2],
Remigijus Didziapetris[1], Pranas Japertas[1]

[1] ACD/Labs, Inc., A.Mickeviciaus g. 29, LT-08117
Vilnius, Lithuania,
[2] Department of Biochemistry and Biophysics,
Vilnius University, M.K.Ciurlionio g. 21/27, LT-03101
Vilnius, Lithuania.

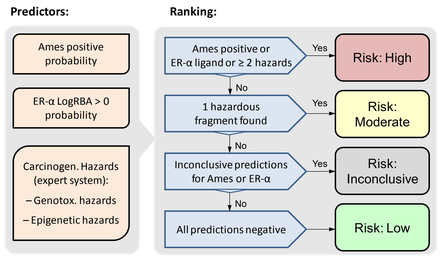
INTRODUCTION

According to FDA Guidance for Industry, assessment of genotoxicity/carcinogenicity by computational methods is sufficient for impurities in drug products present at levels below the ICH qualification thresholds. This study presents a novel approach to aid this assessment based on probabilistic predictors of mutagenicity in Ames test and binding to Estrogen Receptor, supplemented by a knowledge-based system of structural alerts. The list of potentially hazardous structural fragments was compiled from various literature sources and refined by analyzing their performance on data from different assays detecting point mutational and/or clastogenic mechanisms of DNA damage (Ames test, in vitro chromosomal aberrations, micronucleus test, mouse lymphoma assay, sister chromatid exchange). Finally, the expert system was tested on the Carcinogenic Potency Database and FDA carcinogenicity data to ensure detection of common non-genotoxic carcinogens. Selected structural alerts achieved >90% sensitivity for recognizing positive compounds in Ames and Chromosomal Aberrations data sets showing that the absence of alerting groups is a reliable criterion for identifying impurities not posing significant genotoxic/carcinogenic risk.

IN SILICO EVALUATION OF CARCINOGENIC RISK

Compounds may exhibit carcinogenic activity by a multitude of mechanisms. Many carcinogens are genotoxic due to causing either point mutations (mutagenic effect), or chromosomal damage (clastogenic or aneugenic effect), while in the other cases carcinogenicity can be mediated by interactions with specific receptors (non-genotoxic or epigenetic mechanisms).

The aim of the current study was to provide a computational tool that would enable reliable selection of compounds without known alerts for carcinogenic activity. For this purpose, the profiling system should demonstrate very high sensitivity towards multiple classes of hazardous chemicals. Therefore, predictive models were derived for several carcinogenicity-related properties. Their outputs are combined according to the "most unfavorable result" principle, i.e. a chemical is considered hazardous if it obtains positive result in at least one model (Scheme 1).



SCHEME 1. An outline of the compound ranking scheme

AMES TEST

The starting sources of the standardized Ames genotoxicity data set were well known databases:
- Chemical Carcinogenesis Research Information (CCRIS)
- Genetic Toxicology Data Bank (GENE-TOX)

The results of Ames genotoxicity assays were collected for several strains of *S. typhimurium* which are most frequently used for testing (TA97, TA98, TA100, TA102, TA104, TA1535, TA1537, TA1538 and also *E. coli* strain WP2 uvrA), with or without metabolic activation. A compound was considered genotoxic if at least one of conducted Ames test results was positive. Final data set contained about 8,600 compounds with standardized Ames genotoxicity values converted to binary format ("1" – Ames positive, "0" – Ames negative).

The predictive models for Ames genotoxicity was built using the recently introduced GALAS (**G**lobal, **A**djusted **L**ocally **A**ccording to **S**imilarity) modeling methodology. [1] Table 1 briefly illustrates the predictive performance of the obtained model on a test set consisting of 1712 compounds, while more details regarding the Ames test model can be found in [2].

Accuracy testing		Calculated probability (P) <0.5	Calculated probability (P) >0.5	Statistical parameters		Accuracy
Test set (RI ≥ 0.3)* 1,483 compounds 86.6% covered	Safe	**392 (26.4%)**	96 (6.5%)	Specificity	80.3%	**89.0%**
	Genotoxic	67 (4.5%)	**928 (62.6%)**	Sensitivity	93.3%	
Test set (RI ≥ 0.5) 1,117 compounds 65.2% covered	Safe	**257 (23.0%)**	51 (4.6%)	Specificity	83.4%	**93.4%**
	Genotoxic	23 (2.0%)	**786 (70.4%)**	Sensitivity	97.2%	

* RI (Reliability Index) is a built-in measure of prediction reliability. Unreliable predictions (RI < 0.3) were not considered in testing.

TABLE 1. Statistical performance of the predictive models for mutagenicity in Ames test

ESTROGEN RECEPTOR BINDING

Tamoxifen, diethylstilbestrol and other anti-estrogens cause endocrine system disruption mediated by their interactions with estrogen receptors (mostly ER-α). The ligands' binding strength is typically evaluated on the basis of their relative binding affinity (RBA) compared to the reference ligand estradiol.

Experimental LogRBA data were collected from multiple literature publications and converted to binary values using LogRBA = 0 as a cutoff value meaning that the compounds are classified as positive if they exhibit at least 1% of the binding strength of estradiol (by convention LogRBA (estradiol) = 2.00).

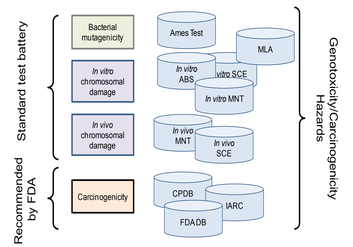
Analogously to Ames mutagenicity, the predictive model for estimating the probability of strong binding to ER-α (LogRBA > 0) was built using GALAS modeling methodology. As shown in Table 2, the model demonstrated very good predictive power: almost 90% of the test set compounds obtained predictions of moderate or high reliability (RI ≥ 0.5).

Accuracy testing		Calculated probability (P) <0.5	Calculated probability (P) >0.5	Statistical parameters		Accuracy
Test set (RI ≥ 0.3) 427 compounds 96.8% covered	LogRBA < 0	**305 (71.4%)**	12 (2.8%)	Specificity	96.2%	**93.7%**
	LogRBA ≥ 0	15 (3.5%)	**95 (22.3%)**	Sensitivity	86.4%	
Test set (RI ≥ 0.5) 389 compounds 88.2% covered	LogRBA < 0	**285 (73.3%)**	9 (2.3%)	Specificity	96.9%	**94.9%**
	LogRBA ≥ 0	11 (2.8%)	**84 (21.6%)**	Sensitivity	88.4%	

TABLE 2. Statistical performance of the predictive model for strong binding to estrogen receptor α.

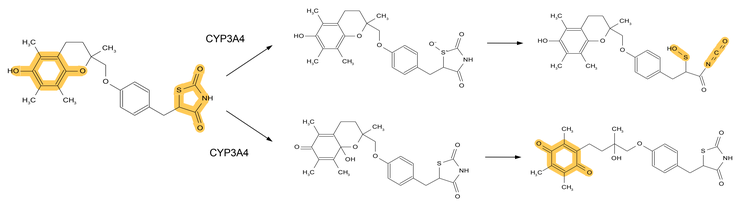
GENOTOXICITY/CARCINOGENICITY HAZARDS

The knowledge-based expert system that identifies structural fragments potentially responsible for genotoxic effect of the compound of interest was derived utilizing experimental data from a variety of assays representing the standard test battery for genotoxicity (Scheme 2). However, none of these assays ensures reliable detection of non-genotoxic carcinogens, whereas according to FDA Guidance the absence of structural alerts for both genotoxicity and carcinogenicity is required for a compound to be adequately qualified by *in silico* methods. Therefore, the expert system was additionally refined on the basis of Carcinogenic Potency database (CPDB), IARC list of human carcinogens, and FDA carcinogenicity data. [3]



SCHEME 2. Assays and databases considered in the development of Hazard identification system

Analysis of these data yielded a list of 67 structural alerts, 14 of which represent epigenetic carcinogens (androgens, peroxisome proliferators, etc.). The alert list is not limited to well-known genotoxic substructures such as planar polycyclic arenes, aromatic amines, quinones, N-nitro and N-nitroso groups, but also includes various other fragments that may undergo biotransformation to reactive intermediates. As an example consider thiazolidinedione class antidiabetic drug troglizatone. It is classified by FDA as a potent carcinogen [3] and has been withdrawn from the USA market. [4] Carcinogenic effect of this drug could be mediated by its several reactive metabolites. In human liver microsomes, chromane ring of troglitazone is metabolized by CYP3A4 to form quinone and quinone-methide products. Furthermore, oxidative cleavage of thiazolidinedione ring results in a reactive sulfenic acid metabolite that also contains an isocyanate moiety. [4,5] As shown in Scheme 3, both bioactivation pathways are predicted by the Hazards identification system presented here. Overall, the Impurity profiler was able to detect 94% of mutagens in the Ames test DB and >90% of compounds that cause chromosomal aberrations *in vitro*.



SCHEME 3. Biotransformation of troglitazone in human liver microsomes.

SOFTWARE FOR PREDICTING GENOTOXICITY/CARCINOGENICITY

The profiling system for impurities and degradants described here is implemented as a part of ACD/Tox Suite 3.0 software package (info@acdlabs.com).

1. Genotoxicity/Carcinogenicity Profile Summary

The Summary Tab accumulates all information provided by separate predictors and calculates the final rank for the analyzed compound (Fig. 1). The output of each particular predictor is categorized ("Good" to "Bad") and the final estimate of carcinogenic risk is given on the basis of the "most unfavorable result" principle:
- **High** – potentially hazardous compound . Experimental testing is needed
- **Moderate** or **Inconclusive** predictions should be evaluated on a case-by-case basis
- **Low** – the compound does not contain known alerts and further testing is not required

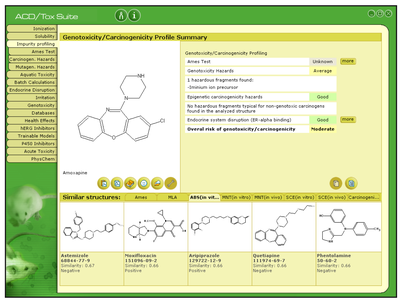


FIGURE 1. Impurity Profile Summary module in ACD/Tox Suite 3.0.

2. Carcinogenicity Hazards

This module presents extended information about the hazardous fragments found in the analyzed molecule, including short description of its mechanism of action, literature references and a bar chart illustrating the distribution of positive and negative compounds in the databases representing different genotoxicity assays. These charts provide further evidence regarding the possible mechanisms of action:
- Mutagens are characterized by high proportions of positive compounds in the Ames test
- Clastogens are better detected in chromosomal damage assays such as ABS or MNT.

E.g., more than 80% of acrylic acid derivatives cause chromosomal aberrations *in vitro*, although the majority of them are Ames negative (Fig. 2). These data suggest that acrylates are primarily reactive towards proteins involved in DNA replication and maintenance, but do not cause direct DNA damage.

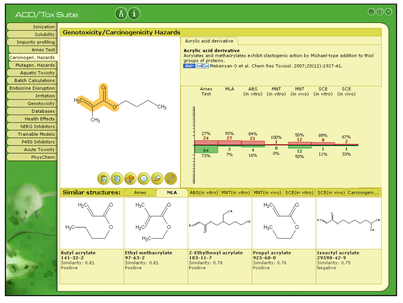


FIGURE 2. Carcinogenicity Hazards module in ACD/Tox Suite 3.0

3. Mutagenicity Hazards

This module provides additional information about hazardous substructures that act by mutagenic mechanisms and are well tested in the Ames test. Here the bar charts represent strain-specific distribution of experimental data and provide insight on types of mutations induced by the particular substructure:
- TA98 strain detects frameshift mutations
- TA100 strain detects base-pair substitutions

Also, comparing the bar charts that display Ames test results with and without metabolic activation allows making a distinction between direct-acting mutagens and compounds that only exhibit hazardous effect after biotransformation.

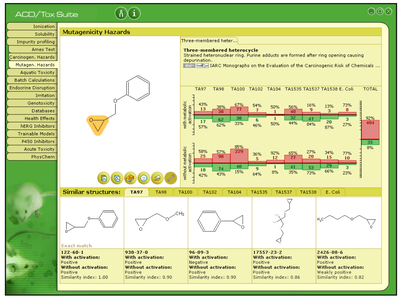


FIGURE 3. Mutagenicity Hazards module in ACD/Tox Suite 3.0

REFERENCES

[1] Sazonovas A et al. *SAR QSAR Environ Res.* **2010,** 21, 127.
[2] Didziapetris R et al. *Toxicol Lett.* **2008,** 180, S152.
[3] The FDA/CDER Carcinogenicity Database.
[4] Kalgutkar AS & Didiuk MT. *Chem Biodivers.* **2009**, 6, 2115.
[5] Mansuy D & Mansette PM. *Arch Biochem Biophys.* **2011**, 507, 1745.

Advanced Chemistry Development, Inc.
Tel: (416) 368-3435
Fax: (416) 368-5596
Toll Free: 1-800-304-3988
Email: info@acdlabs.com
www.acdlabs.com